

APPENDIX A

PRODUCTION FORMAT PROTOCOL

I. PRODUCTION OF ELECTRONICALLY STORED INFORMATION (ESI)

- A. **Load files.** Except where noted in section (L) below, all ESI is to be produced in electronic format, with file suitable for loading into a compatible litigation support review database. All productions will include both image and metadata load files, as described below in Paragraph III, Load File Format.
- B. **Metadata Fields and Processing.** Each of the metadata and coding fields set forth in Paragraph IV that can be extracted from a document shall be produced for that document. The Parties are not obligated to populate manually any of the fields in Paragraph IV if such fields cannot be extracted from a document, with the exception of the CUSTODIAN, PRODVOLID, and TIMEZONE, which shall be populated by the Producing party.
- C. **System Files.** Common system and program files need not be processed, reviewed or produced. The Producing party shall keep an inventory of the system files not being produced and the criteria (*e.g.*, non-human readable file, etc.) for not processing the files.
- D. **Email.** Email shall be collected in a manner that maintains reliable email metadata and structure. Whenever possible, email shall be collected from the Producing Party's email store or server. Metadata and "header fields" shall be extracted from email messages.
- E. **De-Duplication.** Removal of duplicate documents shall only be done on exact duplicate documents (based on MD5 or SHA- 1 hash values at the document level) *across all custodians* (global). The Parties shall retain any metadata that would be lost following deduplication (*e.g.*, Custodian) in a field that lists each distinct value separated by a semicolon. If a Party is unable to maintain such information or if global deduplication could otherwise limit the ability to provide that a particular document was possessed by a custodian, then removal of duplicate documents shall only be done on exact duplicate documents (based on MD5 or SHA- 1 hash values at the document level) *within a source* (Custodian).
- F. **TIFFs.** Single-page Group IV TIFF images shall be provided using at least 300 DPI print setting. Each image shall have a unique file name, which is the Bates/control number of the document. Original document orientation shall be maintained (*i.e.*, portrait to portrait and landscape to landscape). TIFFs will show any and all text and images which would be visible to the reader using the native software that created the document. Documents containing color need not be produced initially in color. However, if an original document contains color necessary to understand the meaning or content of the document, the Producing Party will honor reasonable requests for a color image of the document.

- G. Embedded Objects.** To the extent reasonably possible, non-image files embedded within documents, such as spreadsheets within a PowerPoint, will be extracted as separate documents and treated like attachments to the document in which they were embedded. Graphic objects embedded within documents or emails, such as logos, signature blocks, and backgrounds may, but need not, be extracted as separate documents.
- H. Compressed Files.** Compression file types (*e.g.*, .CAB, .GZ, .TAR, .Z, .ZIP) shall be decompressed in a manner that ensures a container within a container is decompressed into the lowest uncompressed element resulting in individual files. The container file itself shall not be produced.
- I. Text Files.** For each document, a single text file shall be provided along with the image files and metadata. The text file name shall be the same as the Bates/control number of the first page of the document. Electronic text must be extracted directly from the native electronic file unless the document was redacted, an image file, or a physical file. In these instances a text file created using OCR will be produced in lieu of extracted text.
- J. Redaction.** Documents may be subject to redaction of information that (i) regards non-responsive Confidential or Highly Confidential Information, or (ii) is subject to the attorney-client privilege or to the work-product immunity. Each such redaction, regardless of size, shall be clearly labeled as either (i) “Redaction – Relevancy” or (ii) “Redaction – Privilege” according to the Protective Order entered in this action. If a file that originates in ESI needs to be redacted before production, the file will be rendered in TIFF, and the TIFF will be redacted and produced. Spreadsheet files that are produced in native format may be redacted electronically at the Producing Party’s discretion. The Producing Party will provide searchable text for those portions of the document that have not been redacted.
- K. Native Files.** Various types of files, including but not limited to spreadsheets, presentation documents, media files, documents with embedded media files, documents with “macros”, etc., lose significant information and meaning when produced as an image. Any files that are produced in native format shall be produced with a Bates-numbered TIFF image slip-sheet stating the document has been produced in native format, as well as all extracted text and applicable metadata set forth in Paragraph IV.
- a) Spreadsheets.** Excel spreadsheets shall be produced as a native document file along with the extracted text and relevant metadata set forth in Paragraph IV for the entire spreadsheet, plus a Bates-numbered TIFF image slip-sheet stating the document has been produced in native format.
- L. Other ESI that is Impractical to Produce in Traditional Formats (aka, Structured Data).** The Parties understand and acknowledge that certain categories of ESI are structurally complex and do not lend themselves to production as native format or other traditional formats. To the extent a response to discovery requires production of discoverable electronic information

contained in a database, the Parties agree to confer to define appropriate parameters for querying the database for discoverable information and generating a report in a reasonably usable and exportable electronic file (*e.g.*, Excel, CSV or SQL format).

- M. Endorsements.** The Producing Party will brand all TIFF images in the lower right-hand corner with the corresponding Bates/control numbers, using a consistent font type and size. The Bates number must not obscure any part of the underlying data. The Producing Party will brand all TIFF images in the lower left-hand corner with all confidentiality designations, as needed, in accordance with confidentiality definitions as agreed to by the Parties.
- N. Exception Report.** The Producing Party shall compile and retain an exception report enumerating any unprocessed or unprocessable documents, their file type and the file location.
- a) Password-Protected Files.** To the extent any produced documents are password-protected, the Producing Party must either unlock the document prior to production or provide passwords in order to allow access by the Receiving Party. If the Producing Party is unable to process a document because of unknown passwords or other encryption that cannot be cracked using reasonably standard means, the Producing Party shall retain a listing of such documents in an exception report.

II. PRODUCTION OF PHYSICALLY STORED INFORMATION (HARD COPY DOCUMENTS)

- A. TIFFs.** Hard copy paper documents shall be scanned as single page, Group IV compression TIFF images using a print setting of at least 300 dots per inch (DPI). Each image shall have a unique file name, which is the Bates/control number of the document. Original document orientation shall be maintained (*i.e.*, portrait to portrait and landscape to landscape).
- B. Metadata Fields.** The following information shall be produced for hard copy documents and provided in the data load file at the same time that the TIFF images and the Optical Character Recognition (OCR)-acquired text files are produced. Each metadata field shall be labeled as listed below:

FIELD NAME	DESCRIPTION	EXAMPLE / FORMAT
PRODBEGBATES	The production Bates number associated with the first page of a document.	ABC0000001
PRODENDBATES	The production Bates number associated with last page of a document.	ABC0000003
PGCOUNT	Total number of pages for a document.	00006
CUSTODIAN	The name of the primary person the files belong to. This field should be populated as last name, first name.	Doe, John
PRODVOLID	Production volume name.	ABC_PROD001

- C. OCR Acquired Text Files.** When subjecting physical documents to an OCR process, the settings of the OCR software shall maximize text quality over process speed. Any settings such as “auto-skewing”, “auto-rotation” and the like should be turned on when documents are run through the process.
- D. Database Load Files/Cross-Reference Files.** Documents shall be provided with (a) a delimited metadata file (.dat or .txt) and (b) an image load file (.opt), as detailed in Paragraph IV.
- E. Unitizing of Documents.** In scanning paper documents, distinct documents shall not be merged into a single record, and single documents shall not be split into multiple records (*i.e.*, paper documents should be logically unitized). In the case of an organized compilation of separate documents—for example, a binder containing several separate documents behind numbered tabs—the document behind each tab should be scanned separately, but the relationship among the documents in the binder should be reflected in proper coding of the beginning and ending

document and attachment fields. The Parties will make their best efforts to unitize documents correctly.

III. REQUESTED LOAD FILE FORMAT FOR ESI

- a) **Delimited Text File:** A delimited text file (.DAT or .CSV) containing the fields listed in Paragraph III should be provided. The delimiters for the file can be Concordance defaults, but defined delimiters are acceptable:

Comma - ASCII character 20 (,)

Quote - ASCII character 254 (¢)

Newline - ASCII character 174 (¤)

- b) **Image Cross-Reference File (Load File):** The Image cross-reference file (.OPT) is a comma delimited file consisting of six fields per line. There must be a line in the cross-reference file for every image in the database. The format for the file is as follows:

ImageID,VolumeLabel,ImageFilePath,DocumentBreak,PageCount

- ImageID: The unique designation used to identify an image. This should be the Bates number of the document.
- VolumeLabel: The name of the volume.
- ImageFilePath: The full path to the image file.
- DocumentBreak: If this field contains the letter "Y," then this is the first page of a document. If this field is blank, then this page is not the first page of a document.
- PageCount: Number of pages in the document.

Sample Data

```
CNTRL00000001,VOL001,\IMAGES001\CNTRL00000001.TIF,Y,,1
CNTRL00000002,VOL001,\IMAGES001\CNTRL00000002.TIF,Y,,2
CNTRL00000003,VOL001,\IMAGES001\CNTRL00000003.TIF,,,,
CNTRL00000004,VOL001,\IMAGES001\CNTRL00000004.TIF,Y,,4
CNTRL00000005,VOL001,\IMAGES001\CNTRL00000005.TIF,,,,
CNTRL00000006,VOL001,\IMAGES001\CNTRL00000006.TIF,,,,
CNTRL00000007,VOL001,\IMAGES001\CNTRL00000007.TIF,,,
```

IV. REQUESTED METADATA FIELDS FOR ESI

FIELD NAME	DESCRIPTION	EXAMPLE / FORMAT
PRODBEGBATES	The production Bates number associated with the first page of a document.	ABC0000001
PRODENDBATES	The production Bates number associated with last page of a document.	ABC0000003
PRODBEGATTACH	The production Bates number associated with the first page of the parent document.	ABC0000001
PRODENDATTACH	The production Bates number associated with the last page of the last attachment in the document family.	ABC0000008
NATIVELINK	The full path to a native copy of a document.	\\natives\001\ABC0000001.htm
PGCOUNT	Total number of pages for a document.	00006
ATTACHCOUNT	Number of attachments within a document family.	0 (Numeric)
FILENAME	The file name of a document.	Document Name.xls
FROM	The name of the person in the FROM field of every email.	John Doe <jdoe@acme.com>
TO	Recipients of the email. Multiple email addresses should be separated by semicolons.	Jane Smith <jsmith@acme.com
CC	Recipients in the cc: field of the email. Multiple email addresses should be separated by semicolons.	Bob Johnson <bjohnson@acme.com>; Sally May <smay@acme.com>
BCC	Recipients in the bcc: field of the email. Multiple email addresses should be separated by semicolons.	John Doe <jdoe@acme.com>
EMAILSUBJECT	Subject of an email.	Re: resume
DATESENT	Date when an email was sent.	MM/DD/YYYY
TIMESENT	Time an email was sent.	HH:MM:SS
DOCAUTHOR	The author of a document from entered metadata.	John Doe
DOCTITLE	The extracted document title for a loose file or attachment.	Resume.docx
DATELASTMOD	The date a document was last modified.	MM/DD/YYYY
TIMELASTMOD	The time the document was last modified.	HH:MM:SS

FIELD NAME	DESCRIPTION	EXAMPLE / FORMAT
CUSTODIAN	The name of the primary person the files belong to. This field should be populated as last name, first name.	Doe, John
CUSTODIANS_ALL	The name of the primary person/source and all other persons/sources the file belonged to after global deduplication. This field should be populated as last name, first name.	Doe, John; Smith, Jane
FILEPATH	The original path/folder of the source document from the primary person/source	Doe, John\email.pst\inbox
FILEPATHS_ALL	The original paths/folders of the source document from the primary person/source and all other persons/sources after global deduplication	Doe, John\c:\my documents; Smith, Jane\d:\important\
FILEEXT	The file extension of a document.	docx
APPLICATION	Type of document by application.	MS Word, MS Excel, etc.
TIMEZONE	The time zone the document was processed in.	PST, CST, EST, etc.
DATECREATED	The date the document was created.	MM/DD/YYYY
TIMECREATED	The time the document was created.	HH:MM:SS
DATERECEIVED	Date an email was received.	MM/DD/YYYY
TIMERECEIVED	Time an email was received.	HH:MM:SS
HASH	The MD5 or SHA Hash value or “de-duplication key” assigned to a document.	9CE469B8DFAD1058C3B1 E745001158EA
PRODVOLID	Production volume name.	ABC_PROD001
TEXTLINK	The path to the full extracted OR OCR text of the document. Text files should be named per control number or Bates number if the document is produced.	\TEXT\ABC000001.txt
ORGANIZATION	Organization from which data came, e.g., parent company, affiliate, party, etc. to be populated.	Acme, Inc.